2017-12

Big Data, Digitization, and Social Change

Johnson, Jeffrey; Denning, Peter; Sousa-Rodrigues, David; Delic, Kemal A.

ACM

Ubiquity Symposium

Big Data

Big Data, Digitization, and Social Change

by Jeffrey Johnson, Peter Denning, David Sousa-Rodrigues, and Kemal A. Delic

Editor's Introduction

We use the term "big data" with the understanding that the real game changer is the connection and digitization of everything. Every portfolio is affected: finance, transport, housing, food, environment, industry, health, welfare, defense, education, science, and more. The authors in this symposium will focus on a few of these areas to exemplify the main ideas and issues.

Ubiquity Symposium

Big Data

Big Data, Digitization, and Social Change

by Jeffrey Johnson, Peter Denning, David Sousa-Rodrigues, and Kemal A. Delic

The term "big data" is something of a misnomer. Every generation of computers since the 1950s has been confronted with problems where data was way too large for the memory and processing power available. This seemed like an inconvenience of the technology that would someday be resolved when the next generation of computers came along. So what is different about big data today? The revolution is happening at the convergence of two trends: the expansion of the internet into billions of computing devices, and the digitization of almost everything. The internet gives us access to vast amounts of data. Digitization creates digital representations for many things once thought to be beyond the reach of computing technology. The result is an explosion of innovation of network-based big data applications and the automation of cognitive tasks. This revolution is introducing what Brynjolfsson and McAfee call the "Second Machine Age." This symposium will examine this revolution from a number of angles.

Our authors will argue the big data challenge is the not technical problem of moving the maximum amount of bits in the minimum amount of time, but the scientific challenge of formulating methods to represent the complex and entangled systems that we must design and manage to run the modern world. This goes beyond "computational analytics" that abstract statistics from large data sets; "data mining" that discovers useful structures in combined data streams; and "visual analytics" that display multisource data as ingenious maps, color-coded images, and multidimensional videos to aid understanding and inform decisionmaking.

Interconnected Digital World

We will use the term big data with the understanding that the real game changer is the connection and digitization of *everything* [1]. We now have the potential for *anything* knowable to be included in the synthesis of information relevant to the design, planning and management of the many entangled and increasingly complex systems of the modern world.

With this interpretation, big data is the sign that everything is changing. Every portfolio is affected: finance, transport, housing, food, environment, industry, health, welfare, defense, education, science, and more. Our authors in this symposium will focus on a few of these areas to exemplify the main ideas and issues.

Not even education**,** the bastion of learning knowledge-based professions, has escaped this revolution. The most obvious is the emergence of massive open online courses (MOOCs) where hundreds or thousands of students worldwide study online courses at no charge. MOOC platforms now collect and analyze every keystroke and gesture of every student, enabling the system to adjust its pace and style to individual learners. Twenty years ago this was not only impossible, it was unthinkable. Another new emerging education technology is the online competency based module (OCBM), which focuses on having the student develop and demonstrate a specific skill set, such as network configuration or management. Employers are beginning to accept certificates from OCBM companies as evidence of skill in lieu of university diplomas.

Twenty years ago, the technological capability for recording entire courses and keeping detailed student records did not exist. That has changed. Our technology can easily handle complex courses and many thousands of student records. The internet can provide high-speed access around the world. Now data analytics can support "personalized" paths through the syllabus material, attuned to individual student learning styles. For example, using digital data from the half million school students studying a mathematics program in France, a relatively small number of preferred paths can be abstracted and that prospective students can be directed to their most suitable study path for them [2].

Science is being revolutionized by digitization. Fifty years ago scientific knowledge was mostly stored on paper as books, scientific journals, conference proceedings, and handwritten correspondence. Individual libraries gave scientists access to local subsets of this information, but searching for literature often involved very slow inter-library loan requests to other libraries for photocopies of articles or to borrow books. Today all new scientific information is online accessible thanks to rapid internet searches. Scientific societies have placed their entire libraries in online digital libraries. Interestingly there is still a digital divide in scientific information: Publications before approximately 1985 were often not originally available in digital form and had to be scanned as images into PDF documents. This is not suitable for their content to be indexed by internet search engines. However these are increasingly being converted to machine-readable form thanks to OCR. Scientific information is now collected into huge databases such as PubMed and the Digital Patent Archives. Even greater changes in

the organisation of science have been envisaged. For example, in 2014 an international federation of over one hundred universities and institutes, the Complex Systems Digital Campus, or CS-DC (http://www.cs-dc.org), was recognized by UNESCO as a Europe–Latin America–Africa UniTwin network to share scientific and educational resources worldwide. This growing federation of institutions aims to enable new ways of doing science using digital resources. It is organized into e-departments and e-laboratories supporting collaborative research and teaching across the scientific domains. This digital approach promises that collaborative science will be conducted at a fraction of the cost of conventional scientific exchange. The CS-DC will enable outstanding researchers in Africa, Latin America, and other less wealthy regions to participate in collaborative international research. The science research enterprise will never be the same.

Cities exemplify the impact of digitization, connectivity, and ubiquitous computing in the current transformation toward smart cities. Transportation systems within and between cities are managed using local data from fixed and moving sensors, including vehicle navigation systems and drivers' mobile phones. Some cities have extensive video surveillance to support the police and emergency services. Internet-hailed taxi services (Uber) and rooming services (Airbnb) are popular among consumers and contested by city officials. Citizens increasingly demand public internet access, for example when waiting for friends at cafés or checking the progress of the next bus. Smart meters optimize our home use of electricity and water, contributing to city-wide and national targets for sustainability. Indeed it has been found some of us change our behavior to keep up with our environmentally friendly neighbors [3]. Increasingly digitization and connectivity are changing the way we live and the way our cities work, at local, national and global scales.

Economy, Society, and Politics

People around the world are looking to computing technology to solve other large-scale problems. The global financial crash for 2008 showed that the world of banking is much more interconnected than was previously realized. Today's turmoil in North Africa and the Middle East creates huge human and political problems as refugees risk their lives in large numbers to reach safety in Europe—which itself is trying to manage many deep problems cross-border movements. Drug abuse is global with complex interactions between supplying nations, consuming nations, and international crime. Water is increasingly becoming a strategic resource and source of conflict. How can nations and humankind manage these complex

problems? In this symposium our authors investigate the extent to which big data is necessary or sufficient for their solutions in the context of global systems.

The stories above hardly exhaust the supply of examples. Personalized health enables individuals to receive treatment tailored to their own genetic makeups and personal histories, rather than one-size-fits-all standardized treatments. In the meantime, health insurance companies make more precise assessments of an individual's risks, with associated targeted premiums. As another example, big data is routinely used in advertising and marketing to find social trends and target individuals for products and services—sometimes welcome and sometimes intrusive or unsettling.

Another important dimension of big data is its impact on politics. For example, the 2012 re-election of President Barack Obama is partly attributed to targeted use of social media [4]. Similarly, Twitter is considered to have been a factor in regime change during the Arab Spring. Today social media is widely used as a tool for terror. At the same time, those fighting terrorism and crime seek greater power of surveillance by legally gaining access to internet traffic and telephone data, thereby compromising the freedoms and privacy of innocent citizens. Big data is a two-edged weapon, serving crime and terror with the same indifference that it serves democracy and freedom. For example, it has recently been suggested Russian hackers subverted the 2016 U.S. presidential election [5].

The Digital Age is Here

The turn of this century was marked by one silent, but very crucial moment: For the first time in human history, the volume of digital data has surpassed the volume of analog data produced. Not only has rapid growth of the social networks and its number of users exploded, but also converging technologies have enabled omnipresent, always connected devices producing torrents of a wide variety of data generated each and every second. Initially it was observed as an interesting millennial-driven phenomena, but it soon became obvious that important economic changes will follow. Right now, we see a number of jobs are in peril thanks to digital content, services, and transactions. Internet shopping evolved into a new economy, called sometimes "service economy" and recently renamed into "digital economy." Interestingly enough, as humanity passed from Agriculture Age into Industrial Age and now enters Digital Age, we see very similar phenomena happening. New technologies are destroying old jobs and are creating new ones, but not with the same speed or quality. Large migrations are provoked by the new economic circumstances, and societies are divided into supporters and opponents

to inevitable technological changes. We observe the each and every age brought technologies with double-sword effect: producing new and destroying old. Thus, we can only hope that big data (as raw material) will bring more benefits than damages to the forthcoming digital economy (digital goods and services) for the benefits of the entire humanity. But inevitable already started: We live in the digital age.

About the Authors

Jeffrey Johnson is Professor of Complexity Science and Design at the Open University in the UK, Deputy President of the UNESCO UniTwin Complex Systems Digital Campus (http://www.cs-dc.org), Past-President of the Complex Systems Society (https://cssociety.org), and a partner in the European Erasmus Da.Re. Project (Data Science Pathways to Re-imagine Education, http://dare-project.eu). His research interest is in the dynamics of complex multilevel social and environmental systems, and in systems thinking and computational complex systems science in policy and management. He has published many scientific books and papers, and supervised a large number of doctoral students. He has many years experience of online education including creating MOOCs on global systems science and policy, systems thinking and complexity, and data science. He is CEO of Vision Scientific Ltd. and had led many scientific and commerical research and development projects. He is an associate editor of ACM *Ubiquity*.

Peter J. Denning is Distinguished Professor of Computer Science and Director of the Cebrowski Institute for information innovation at the Naval Postgraduate School in Monterey, California, is Editor of ACM *Ubiquity*, and is a past president of ACM. The author's views expressed here are not necessarily those of his employer or the U.S. federal government.

David Sousa-Rodrigues is a professor of Computer Science, Complexity and Design at the European University (https://www.europeia.pt/), in Lisbon, and at Polytechnic Arts and Design School (http://www.esad.ipleiria.pt/) in Caldas da Rainha, Portugal. He is a member of the Centre for Complexity and Design research group of the Open University in the U.K., and was member of Complex Systems Society permanent council. His research interests are distributed multilevel complex systems and bio-inspired computational modeling. Specific topics of interest include social networks, evolutive computation, machine learning, data analytics, community detection, cellular automata, emergence, and stigmergy.

Kemal A. Delic is a senior technologist and practicing enterprise architect with Hewlett-Packard Co. He is Visiting Senior Fellow with The Open University, UK. He is also adjunct professor at IAE

business school at Grenoble University. He serves as associate editor to ACM *Ubiquity* magazine. He acted as an advisor and consultant to European Commission on FET programs. He holds two U.S. and one E.U. patent. Delic lives in Grenoble, France. He holds a Dipl. El. Ing. degree from the University of Sarajevo. During the last 30-plus years he has worked mainly in the area of very large-scale systems—either on industrial, commercial products or academic research. His principal interest is in architecture, design, and engineering of large-scale systems. From 1989 to 1990, and again from 1994 to 1996, he was a visiting professor and researcher with CNR – Italian National Research Council – IEI institute working on error detection through signature analysis (PDCS Project), Bayesian Belief Network models in safety arguments (EU SHIP Project), and marketing with mobile intelligent agents (EU MIA Project). He has published 100-plus papers, articles, and essays in journals, magazines, and conferences. He has given talks, delivered lectures, and organized workshops and conferences. In the 1980s he published original book on Pattern Recognition Principles and more recently contributed three book chapters on "Enterprise Knowledge Clouds" (2010, 2011). His recent research interest is in the science and practice of hybrid complex systems. Big science, big data and the digital economy are his most recent interests.

DOI: 10.1145/3158335

References

[1] Brynjolfsson, E., McAfee, A. *The Second Machine Age*. W.W. Norton and Co., New York, 2014; https://tanguduavinash.files.wordpress.com/2014/02/the-second-machine-age-erik-brynjolfsson2.pdf

[2] Valigiani, G., Lutton, E., Collet, P. Adapting the ELO rating system to competing subpopulations in a "man-hill." In *Proceedings of the 13th {ISPE} International Conference on Concurrent Engineering (Sept. 18-22, Antibes, France).* IOS Press Amsterdam, The Netherlands, 2006, 766–773.

[3] Behaviour change and energy use. Cabinet Office Behavioural Insights Team. United Kingdom. July 2011. Ref: 406537/0711; https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/48123/2135-behaviour-change-and-energy-use.pdf

[4] Issenberg, S. How President Obama's campaign used big data to rally individual voters. *MIT Technology Review.* Dec. 19, 2012; http://www.technologyreview.com/featuredstory/509026/how-obamas-team-used-big-data-to-rally-voters/

[5] Harris, S .Congressional calls grow for a probe of Russian hacking in U.S. election. *Wall Street Journal*. Dec. 12, 2016; http://www.wsj.com/articles/congressional-calls-grow-for-a-probe-of-russian-hacking-in-u-s-election-1481151191